# A helix scaffold for the assembly of active protein kinases

**Alexandr P. Kornev*, Susan S. Taylor*†‡, and Lynn F. Ten Eyck†§**

*Howard Hughes Medical Institute, †Department of Chemistry and Biochemistry, and §San Diego Supercomputer Center, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093

Contributed by Susan S. Taylor, August 12, 2008 (sent for review July 18, 2008)

**Structures of set of serine-threonine and tyrosine kinases were investigated by the recently developed bioinformatics tool Local Spatial Patterns (LSP) alignment. We report a set of conserved motifs comprised mostly of hydrophobic residues. These residues are scattered throughout the protein sequence and thus were not previously detected by traditional methods. These motifs traverse the conserved protein kinase core and play integrating and regulatory roles. They are anchored to the F-helix, which acts as an organizing "hub" providing precise positioning of the key catalytic and regulatory elements. Consideration of these discovered structures helps to explain previously inexplicable results.**

graph theory | hydrophobic motifs | structure comparison

Protein kinases represent a large protein superfamily that regulates numerous processes in living cells (1). Malfunction of this regulation typically leads to various diseases, including immunodeficiencies, cancers, and endocrine disorders (2). Multiple sequence alignment identified the most conserved motifs and defined universal subdomains in protein kinases (3). Solving crystal structures of different protein kinases demonstrated not only a conserved core but also the exceptional flexibility of protein kinases. This indicated an important role of dynamics and plasticity for this family (4, 5). Substantial progress has been made in understanding the regulatory mechanisms, although many questions still remain unanswered (6). Recently, we reported a new bioinformatics method that is capable of detecting conserved patterns formed by residues in space without any relation to protein sequence or main chain geometry. Originally, it was created for comparison of protein surfaces (7, 8), but later the method was expanded for analysis of the whole molecule and was termed "Local Spatial Patterns (LSP) alignment" (9). Application of the method to a set of serine/threonine and tyrosine kinases led to the discovery of an unusual structure, which we termed a "spine" (8). The most remarkable feature of the spine is that it is assembled during the protein kinase activation process and provides coordinated movement of the two kinase lobes. In deactivated kinases, the spine is usually broken because of the rearrangement of the C-helix and/or activation loop. Disassembly of the spine leads to general destabilization of the kinase molecule, which was previously observed in hydrogen–deuterium exchange studies (10, 11) and MD simulations (12). It was demonstrated subsequently that mutation of the spine residues leads to increased flexibility of the activation loop in MAP kinase ERK2 (13) and to a total inactivation of p38$\alpha$ MAP kinase (14).

Despite the fact that the spine is a conserved feature, present in all active eukaryotic protein kinases, it was not detected earlier as a conserved spatial motif. This is due, in part, to the highly unusual nature of its formation. It is comprised of four single residues coming from four different subdomains of the protein kinase molecule (III, IV, VIb and VII)¶, which do not form a sequence "motif" in a traditional sense. 3D alignment of different kinases was also incapable of detecting this ensemble, because it does not form any contiguous main-chain pattern. This discovery drew our attention to the internal structure of the kinase catalytic core. Quite often, it is considered as an amorphous clustering of hydrophobic residues, a result of hydrophobic collapse in the protein folding process. However, it was shown that large ensembles of residues can be formed inside proteins, thereby creating allosteric signaling pathways (15, 16). Residues in these formations are precisely positioned, and their mutation abolishes the allosteric signal propagation. Existence of the hydrophobic spine demonstrated that such unconventional structures not only may be a part of allosteric signaling systems but also can perform structural and/or regulatory functions. Detection of such ensembles is not a trivial task, because they can be formed by residues that come from different parts of the polypeptide chain and do not form any motifs in terms of sequence or secondary/tertiary structure. Usually it requires a complicated multiple sequence alignment of hundreds or even thousands of sequences to achieve statistical equilibrium (15). In contrast, the LSP alignment does not need any sequence alignment, although it does require knowledge of the 3D structures. Its advantage, however, is that comparison of only two structures may provide meaningful insight into protein structure and function. The reason is that the LSP alignment not only detects the conserved patterns of residues but also ranks these residues according to their involvement in the patterns.

In our previous work, we analyzed only water-accessible residues (8). In this study, we used the LSP alignment to compare whole molecules of serine/threonine and tyrosine kinases. Inside the hydrophobic core of the kinases, we detected conserved unconventional motifs, similar to the spine reported earlier. We show that the F-helix, which is positioned in the middle of the large lobe, plays an integrating role by anchoring many hydrophobic motifs. These motifs traverse the entire molecule and orchestrate the catalytic process. They also provide diverse mechanisms for regulation. We define a second "spine," which we term a catalytic spine, because it traverses both lobes but is completed by the adenine ring of ATP. This is distinct from the previous spine that we now refer to as the regulatory spine. The R and C spines are anchored to the N and C termini of the F-helix, respectively. We furthermore demonstrate that the APE motif, which is bound to the F-helix, nucleates substrate binding and allostery. A discussion of several published works shows that newly defined structures can be helpful for understanding the intramolecular machinery of protein kinases and their interactions with other proteins.

Author contributions: A.P.K., S.S.T., and L.F.T.E. designed research; A.P.K. performed research; A.P.K. contributed new reagents/analytical tools; A.P.K., S.S.T., and L.F.T.E. analyzed data; and A.P.K. and S.S.T. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

‡To whom correspondence should be addressed. E-mail: staylor@ucsd.edu.

¶Subdomains are numbered according to Hanks and Hunter (3).

This article contains supporting information online at www.pnas.org/cgi/content/full/0807988105/DCSupplemental.

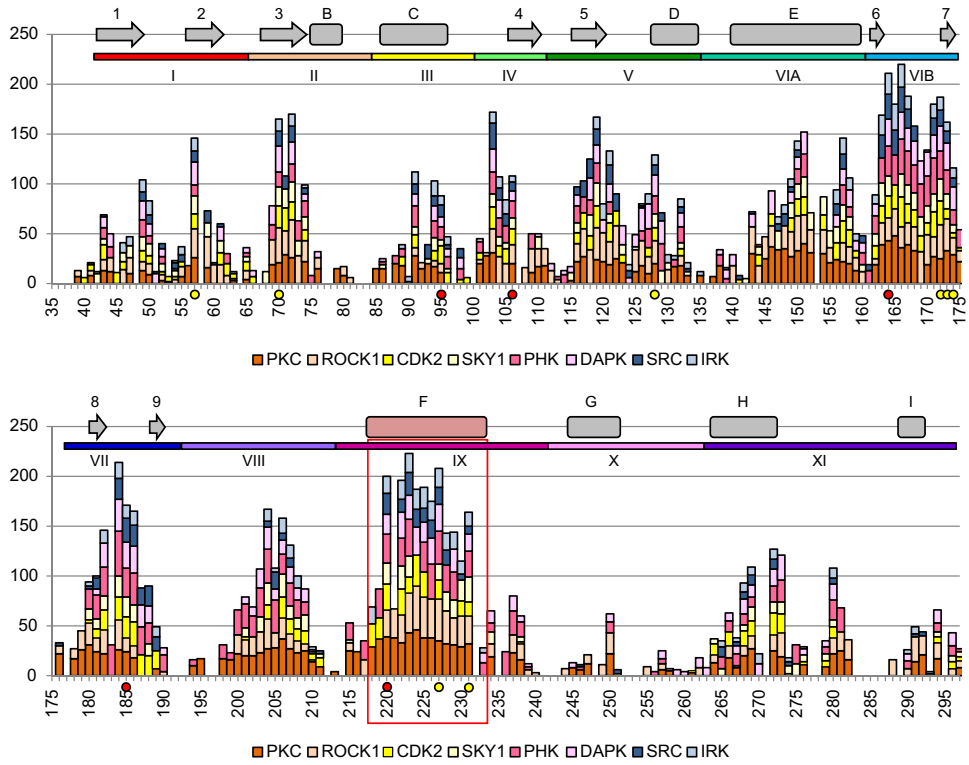


**Fig. 1.** LSP alignment of active conformation of PKA to active conformations of other kinase. Four different families were used in the comparison: AGC kinases (PKC and ROCK1), CMGC kinases (CDK2 and SKY1), calcium/calmodulin kinases (PHK and DAPK), and tyrosine kinases (SRC and IRK). Residues that constitute the two conserved spines are spread along the kinase sequence: R spine residues are marked as red dots; yellow dots mark the C spine residues. Highly scored αF-helix is marked by red square.

## Results

**LSP Alignment of Different Serine/Threonine and Tyrosine Kinases.** The most important information produced by the LSP alignment is expressed by the so-called involvement score (IS). It is defined for every residue of the compared proteins. If the score equals zero, this residue is not a part of any conserved spatial pattern. The higher the IS value, the more this residue is involved in formation of the conserved patterns, and the higher the probability that it is important for protein functionality [see supporting information (SI) *Text* for detailed explanation of the concept]. Although in the previous work (8), we described the major changes related to protein kinase activation, in this work, we have analyzed only active kinases and considered all residues, not just solvent accessible residues, as was done previously. We compared protein kinase A (PKA) from the AGC group to eight kinases from four groups: AGC (PKC and ROCK1), CMGC (CDK2 and SKY1), CaMK group (PHK and DAPK), and protein tyrosine kinases (SRC and IRK). All IS values obtained in eight comparisons are listed in the Table S1. Fig. 1 shows accumulated IS for each PKA residue.

Newly detected residues with high IS were localized mostly in the hydrophobic core of the large lobe, predominantly in helices E and F, around the $P$ + 1-loop and the APE motif (residues 201–210; here and subsequently, we will use mammalian PKA sequence numbering) (Fig. 1, Table S1). Several residues scored in every kinase were found in the H-helix. In addition, residues with exceptionally high IS, which were not appreciated previously, were detected in the N lobe and catalytic and activation loops: $L^{103}$, $M^{128}$, $I^{150}$, $L^{172}$, and $V^{182}$.

**Defining Conserved Structural Motifs.** Traditionally conserved motifs are related to an amino acid sequence pattern (like DFG or APE motifs in protein kinases) or a combination of secondary structures (e.g., helix–turn–helix motif in DNA-binding proteins). However, as we showed earlier (8), proteins can contain unconventional structural motifs, which are not related to any sequence or secondary structure motifs. This raises important questions: what constitutes a conserved structural motif? Where is a border between two motifs? In the case of traditional motifs, the answer to the second question is intuitive; two different motifs have to be separated "far enough" from each other. However, in our case, when motifs are formed by amino acid residues coming from different parts of the primary sequence, this approach is irrelevant. In the current work, we have attempted to separate the large set of highly conserved residues, identified by the LSP alignment, into several different structural motifs based on their catalytic or regulatory roles.

For several reasons, we considered the F-helix as the central hub for these motifs. First, the IS level obtained by the F-helix residues was the highest through the entire molecule. Only the catalytic loop region together with the $\beta_6$ and $\beta_7$ strands was comparable. Second, it is positioned in the middle of the rigid hydrophobic core of the large C lobe, which means it is one of the most immobilized secondary structures in the molecule. Third, it is known that the F-helix serves as a signal-integrating element, which connects several key areas such as the substrate-binding residues and the catalytic loop (5, 17). These observations define the F-helix as a robust scaffold for the whole protein kinase molecule, where all of the motifs can be precisely positioned in space with respect to each other.

**R Spine.** Earlier, we defined the regulatory spine as a motif of four hydrophobic residues, connecting the two lobes of the kinase (8). In the current work, we demonstrate that all structural motifs known to be important for catalysis are connected to the F-helix. In the case of the hydrophobic regulatory spine, this role is

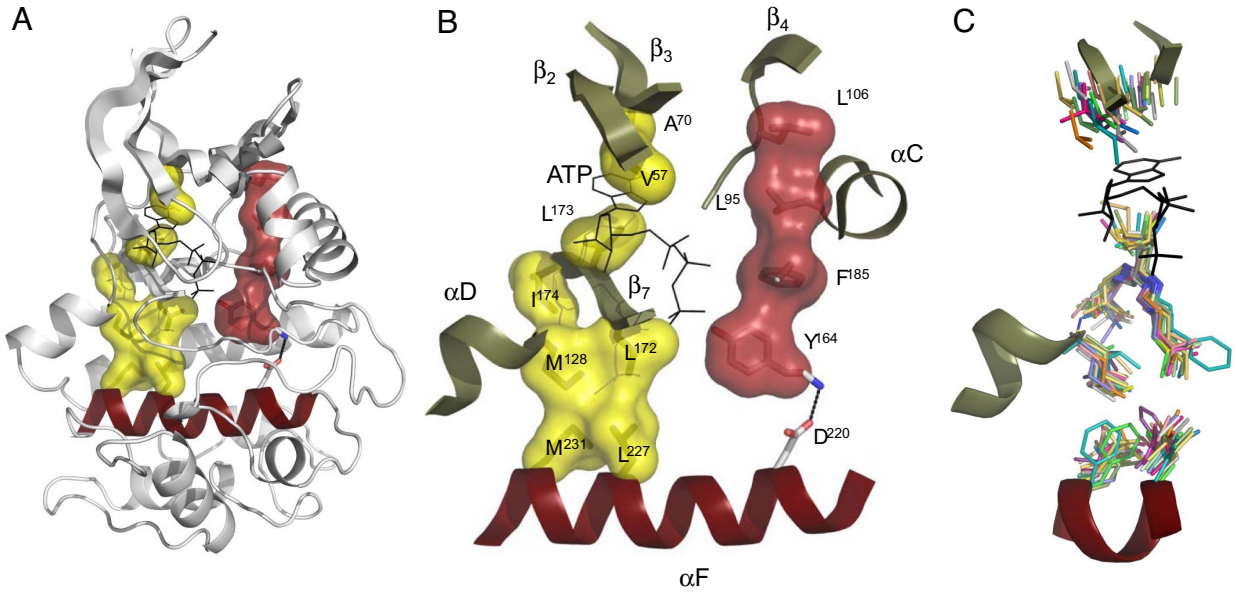


**Fig. 2.** R and C spines flank the F-helix and span the protein kinase core. The PKA structure is shown as a prototype. (*A* and *B*) The hydrophobic part of the R spine is shown as a red molecular surface. The C spine is colored yellow. The adenine ring of ATP completes the C spine. (*C*) Eight hydrophobic residues that constitute the C spine in 22 active kinases from six different groups (see Table S2 for the list of kinases).

played by the aspartic acid $D^{220}$ (Fig. 2), which is universally conserved in all eukaryotic and eukaryotic-like kinases (5, 17) and is the first F-helix residue with high IS. Further, we will refer to this five-residue motif ($L^{95}$, $L^{106}$, $Y^{164}$, $F^{185}$, and $D^{220}$) as regulatory spine, or R spine.

**C Spine.** The last highly scored residue at the C terminus of the F-helix was $M^{231}$ (Fig. 1). To the best of our knowledge, this residue was never considered as an important part of the kinase structure. However, our analysis shows that it starts a highly scored hydrophobic structure, which spans the large lobe and contacts the adenine moiety of ATP (Fig. 2). This structure includes two residues from the F-helix: $L^{227}$and $M^{231}$; one residue from the D-helix, $M^{128}$; and the three hydrophobic residues, which comprise the $\beta_7$-strand: $L^{172}$, $L^{173}$ and $I^{174}$. There are two highly ranked residues that contact the other side of the adenine ring: $V^{57}$ and $A^{70}$. Thus, in the presence of ATP, this ensemble connects the two lobes together, in a way similar to the R spine. By analogy, we termed this motif catalytic spine, or C spine. The major difference between these two spines is that the R spine formation depends on activation loop configuration, whereas the C spine depends on the presence of ATP. The C spine thus defines the adenine ring as a primary driving force for integration of the small and large lobes. As shown on the Fig. 2*C*, the C spine is conserved in both serine/threonine and tyrosine kinases. However, in more distant kinases, such as 3′,5′′-aminoglycoside phosphotransferase (APH) (18), this part of the molecule has a substantially different organization. It should be noted that the R spine, unlike the C spine, is present in APH (8). This suggests that the C spine is more sensitive to the nature of the substrate to be phosphorylated, whereas the R spine remains conserved in different kinases as soon as they have the conserved HRD and DFG motifs and the regulatory C-helix.

**H-Helix Anchor.** There are three hydrophobic residues ($W^{222}$, $I^{228}$, and $Y^{229}$) and glycine ($G^{225}$) from the middle part of the F-helix, which scored very high. These residues form a tight hydrophobic cluster with the four conserved leucines in the H-helix ($L^{268}$, $L^{269}$, $L^{272}$, and $L^{273}$). As Fig. 3 shows, other residues from previously described structures are involved in the cluster: $W^{222}$, $V^{226}$, and $I^{228}$. An apparent function of the structure is to protect and stabilize the hydrophobic core around the F-helix. A characteristic feature of the structure is the position of $L^{273}$, which is inserted into the void space between $W^{222}$ and $V^{229}$, where a strictly conserved $G^{225}$ is located. Such insertion is a characteristic of tetratricopeptide repeats (19), which provide a robust connection between two antiparallel $\alpha$-helices. Most likely, the strong conservation of $G^{225}$ is dictated by this feature.

**Catalytic and Activation Loops Anchoring.** The maximal involvement score in our computation was received by another residue from the F-helix: $A^{223}$. It was slightly higher than IS for such catalytically crucial aspartates as $D^{166}$ and $D^{184}$ (Fig. 1, Table S1). As shown in Fig. 4 *A* and *B*, $A^{223}$ and neighboring $L^{224}$ perform a very important function of anchoring $L^{167}$ from the catalytic loop. This leucine interacts both with R and C spines and provides exact positioning of $D^{166}$ and $K^{168}$ main chain (Fig. 4*A*). These residues are directly involved in phosphoryl transfer (20) and are universally conserved.

$L^{224}$ also serves as a basis for another highly scored residue from the E-helix: $I^{150}$ that defines the position of N terminus of the activation loop with the $D^{184}$ (Fig. 4 *B* and *C*). Stabilization of this aspartate is critical, because it binds a magnesium atom that bridges the $\beta$- and $\gamma$-phosphates of ATP. Such positioning is secured by a contact between $I^{150}$ and two hydrophobic residues from the $\beta_8$ strand: $I^{180}$ and $V^{182}$. They interact with both spines and are a part of the $\beta_7$-$\beta_8$ sheet, a structural feature that is conserved in all protein kinases.

**Substrate-Binding Structure.** A common feature for all protein kinases is that they phosphorylate polypeptide chains. This separates them from all other phosphotransferases and makes their substrate-binding structure substantially different (21). This structure is known to be formed by two major structural elements: the G-helix and the C-terminal part of the activation segment (5), referred to as the *P* + 1 loop (residues 198–205), where the substrate is bound to provide positioning of the phosphorylation site next to the $\gamma$-phosphate of ATP. Our analysis identified a set of highly scored residues in the subdomain VIII, with the conserved APE motif at the end of the

BIOCHEMISTRY

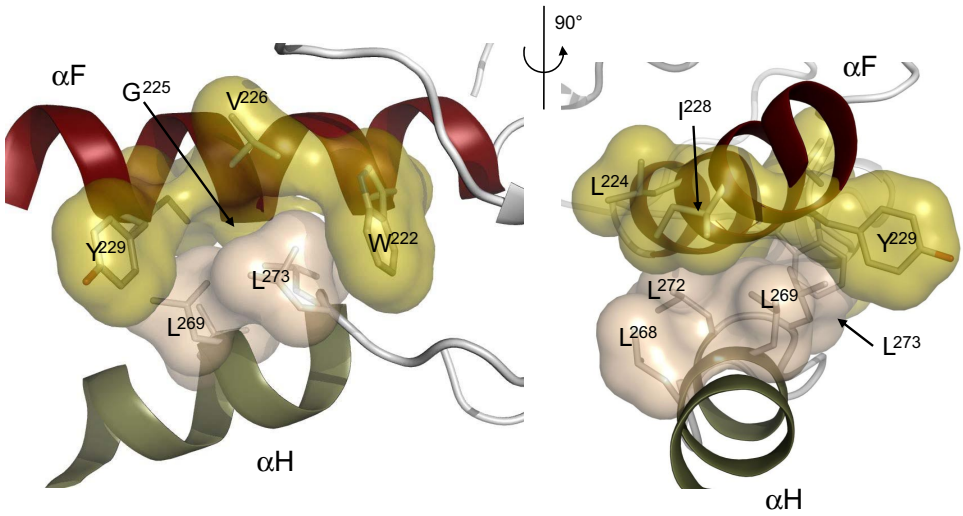


**Fig. 3.** αH-helix anchoring structure in PKA. Molecular surfaces of the highly scored residues are shown: olive, from the αF-helix, tan, from the αH-helix. $L^{273}$ fits as a ''knob'' into a ''hole'' formed by conserved $W^{222}$, $G^{225}$, and $Y^{229}$.

activation segment. Residues from $Y^{204}$ through $E^{208}$ were detected in both serine/threonine and tyrosine kinases (Fig. 1, Table S1). Several residues received high scores only in serine/threonine kinases: $T^{201}$, $E^{203}$, and $I^{209}$. The major difference between serine/threonine and tyrosine kinases is that the size of tyrosine side chain is significantly larger. This explains the variation in geometry and sequence of the $P + 1$ Loop, where a hydrophobic residue following the phosphorylation site is accommodated (Fig. 5*B*).

The highest levels of IS were obtained by $Y^{204}$ and $A^{206}$, which are in direct contact with the highly scored residues from the F-helix: $W^{222}$ and $V^{226}$ (Fig. 5*B*). This is consistent with our suggestion that all of the important structural motifs are connected to the F-helix. As we showed earlier, these two residues from the F-helix constitute a part of the rigid H-helix anchoring structure. Fig. 5 shows that they also play an important role as a foundation of the substrate-binding structure. $V^{226}$ provides the exact positioning of the conserved tyrosine ($Y^{204}$) in serine/threonine kinases (Fig. 5*A*) or tryptophan in tyrosine kinases. This tyrosine was reported to be an important allosteric ''hot spot'' in PKA (22–24). It is involved in hydrophobic interaction with the aliphatic part of $K^{168}$ from the catalytic loop and, together with another residue from the $P + 1$ loop ($T^{201}$), stabilizes the side chain of this catalytically important residue. $W^{222}$ tightly binds to the Ala-Pro pair from the APE-motif (Fig. 5 *B* and *C*) and nucleates the hydrophobic structure that positions four hydrophobic residues comprising the $P + 1$ loop (Fig. 5*B*). The APE-glutamate ($E^{208}$) anchors the side chain of the highly conserved arginine from the loop between the H and I helices: $R^{280}$ (Fig. 5*C*).

## Discussion

During last two decades, protein kinases were studied intensively using sequence and structure alignments. Our recent work demonstrated that comparison of protein kinase surfaces can discover unconventional structural motifs that play an important regulatory role (8). Further development of the surface comparison method allowed us to include both water-exposed and buried residues in the analysis (9). The major objective of this work is to analyze conserved structural features in the hydrophobic core of different protein kinases. As we expected, all residues, which were known to be important for catalysis or for its regulation were highly scored, thus confirming the credibility of our method. Along with the well known residues, we discov-

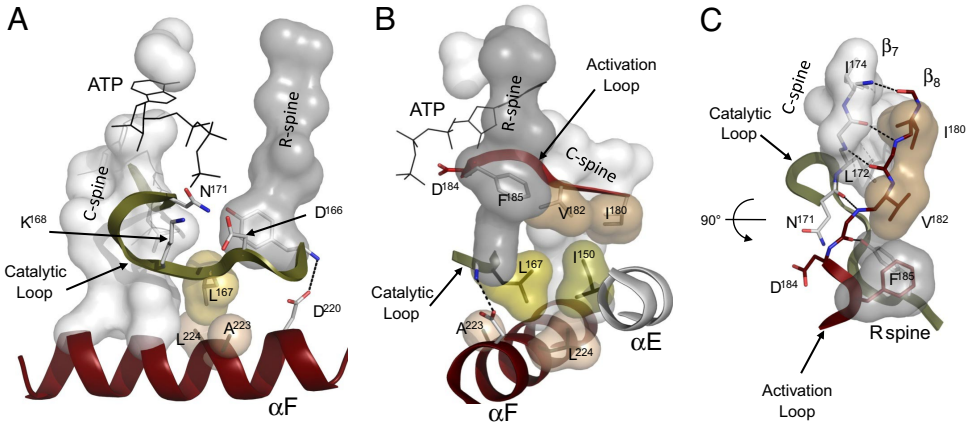


**Fig. 4.** Catalytic and activation loop anchoring in PKA. (*A*) N- and C-terminal parts of the catalytic loop are secured by the spines. The middle part is docked to $A^{223}$ and $L^{224}$ from the αF-helix. Three catalytically active residues are shown: $D^{166}$, $K^{168}$, and $N^{171}$. (*B* and *C*) Activation loop anchoring. Its N terminus binds to the $\beta_8$-strand that makes both polar and hydrophobic bonds to the $\beta_7$-strand from the C spine. The $\beta_8$-strand is also bound to the αF-helix via hydrophobic interaction with the conserved $I^{150}$. This positions the N terminus of catalytically active D184, whereas its C terminus is secured by the R spine.

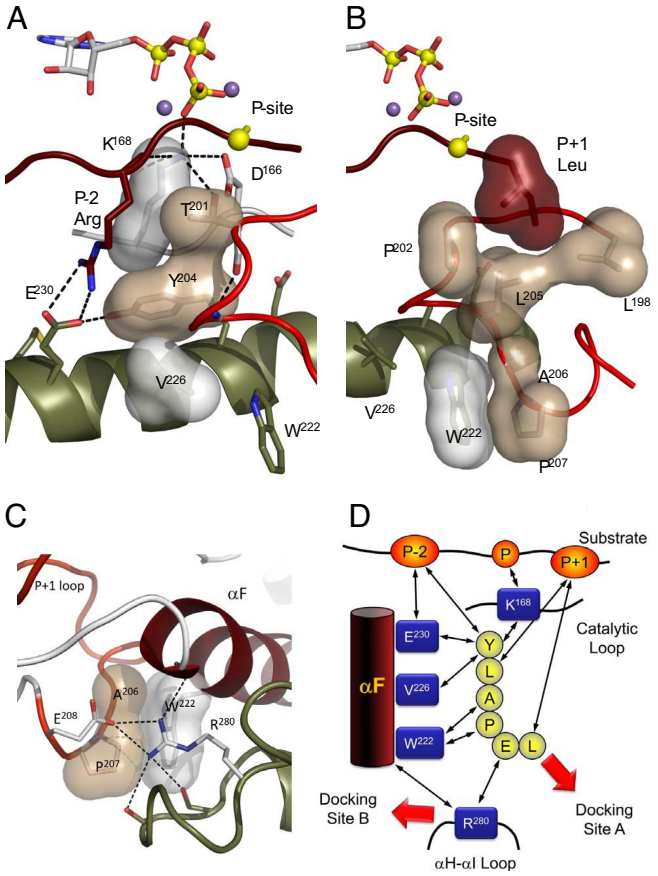


**Fig. 5.** Organization of the substrate-binding network in PKA. (*A*) $V^{226}$ stabilizes $Y^{204}$, which anchors the activation segment (colored bright red) and orients side chain of $K^{168}$. $E^{230}$ interacts with the substrate arginine. (*B*) $W^{222}$ serves as a docking platform for the APE motif, which serves as a foundation for the *P* + 1 loop. (*C*) Closeup of the APE-motif interactions. $R^{280}$ forms multiple hydrogen bonds to the αF-helix, the APE motif, and the αH-αI loop. (*D*) General chart of the YLAPEL motif that mediates interaction between the αF-helix, substrate binding, and docking sites.

ered a set of residues that were not considered previously to be important elements of the protein kinase structure or function (Table S3). The major problem, however, was to separate the pool of the detected residues into meaningful structural elements. In the analysis, we followed a simple principle of functionality: there has to be a structure that coordinates the binding of ATP and a structure that coordinates the binding of the substrate. These structures have to be positioned with respect to each other in a precise and robust way. In our analysis, we found that the highest levels of the IS were obtained by the F-helix. Almost every residue of the helix was highly conserved and precisely positioned with respect to the other key residues responsible for catalysis, substrate binding, or regulation. Earlier, the helix was described as a signal integrating motif for the protein kinase molecule (5). In this work, we demonstrate that it serves as a central scaffold for active protein kinase assembly. Two conserved structures that flank the F-helix and span the protein kinase molecule are the most significant feature of the intramolecular architecture (Fig. 2). We termed these structures R and C spines. They represent unconventional structural motifs formed by residues coming from different parts of the protein sequence and do not form any sequential motifs in a traditional sense (Fig. 1). An important property of these structures is that they can be assembled and disassembled depending on the presence of ATP, the phosphorylation state of the activation loop, or the C-helix movement. This provides a significant conformational plasticity, which is very important for protein kinase functionality (4, 5). Furthermore, because the spines are comprised of hydrophobic residues, this makes the connection between different parts of the molecule firm but flexible. When the two spines are assembled, all catalytically important residues are securely positioned and ready for the catalysis (Fig. 4).

Substrate binding is also defined by residues from the F-helix: $W^{222}$, $V^{226}$ and $E^{230}$ (Fig. 5). The latter directly binds to the substrate and, for this reason, is relatively substrate specific. It was shown that presence of glutamate in this position is a characteristic of kinases with substrate preference for P-2 or P-5 arginines (25). These three residues position the C-terminal portion of the activation segment, which is usually associated with the APE-motif. Fig. 5*D* shows the general organization of the substrate binding elements. $E^{230}$, together with $V^{226}$, anchors $Y^{204}$, which is highly conserved in serine/threonine kinases but substituted by tryptophan in tyrosine kinases. However, both tyrosine and tryptophan play a common role by providing a rigid platform for the N-terminal part of the substrate and by orienting the side chain of the catalytically active $K^{168}$ (arginine in tyrosine kinases) (Fig. 5*A*). $W^{222}$ serves as a docking site for binding alanine and proline in the conserved APE-motif (Fig. 5 *B* and *D*). The hydrophobic residue right before the alanine ($L^{205}$ in PKA) defines the position of the *P* + 1 loop, which stabilizes the C-terminal part of the substrate (Fig. 5*B*). The main chain geometry of the six C-terminal residues of the activation segment is universally conserved through all serine/threonine and tyrosine kinases, even in casein kinase I (PDBID 2CHL), which does not have the APE-motif, $W^{222}$ or $R^{280}$. This geometry is also precisely mimicked in the transphosphorylating protein kinases, when two interacting kinases exchange their activation segments (26). It emphasizes the crucial role of the YLAPEL-motif. Fig. 5*D* shows that it is related not only to substrate binding but also to docking sites on the protein kinase. Docking site A is located between the activation segment and the G-helix. It is a major binding site between the catalytic subunit of PKA and the A-domain of its regulatory subunits (27, 28). It was shown that docking of the CDK-interacting protein phosphatase KAP to CDK2 also occurs on docking site A (29). The second site is located in the loop between the H and I helices. This was shown to be an important interface in the PKA type RII holoenzyme, where the B-domain of the regulatory subunit binds to the catalytic subunit (28). One can suggest that, because of the proximity of this area to the substrate binding site and the universally conserved interaction between the APE-glutamate and $R^{280}$, docking site B is important for other protein kinases as well.

Thus, we defined several conserved spatial structures in the hydrophobic core of protein kinases, which position the following three major elements: the peptide substrate, ATP, and the set of residues directly involved in the catalysis. Definition of the structures and their roles helps to interpret previously inexplicable results. For example, it was not clear why the $E^{230}Q$ mutant of PKA crystallized in an open apo-form, despite the fact that ATP was present in the solution (30). $E^{230}$ is usually considered a part of substrate-binding complex, and it was not clear how it could influence ATP binding. However, a close look at the $E^{230}Q$ mutant structure shows that such a glutamate–glutamine substitute significantly destabilizes $R^{133}$ from the D-helix, which is bonded to the $E^{230}$ in wild-type PKA. The arginine side chain in the $E^{230}Q$ mutant was not resolved, and the whole D-helix had elevated temperature factors. According to our classification, the D-helix constitutes the middle part of the C spine, which is responsible, in particular, for ATP binding (Fig. 2). Destabilization of the C spine foundation would explain the observed instability of ATP binding. This would lead to inability of the mutant to complete the C spine formation, promoting the apo-configuration.

A similar effect of the C spine destabilization was observed in the $F^{314}A$ mutant of PKA (31). This phenylalanine is located in the C-terminal tail and contacts $H^{131}$ and $I^{135}$ from the D-helix. Clearly, the phenylalanine-to-alanine mutation had to destabilize both the C-terminal tail and the C spine. Indeed, this mutation led to a significant decrease in thermal stability, a moderate decrease in affinity for ATP, and a nearly 20-fold decrease in the catalytic activity. Mutation of the neighboring residue $I^{315}$, which is not in contact with the D-helix, and thus the C spine, decreased thermal stability of the mutant and affinity for ATP and Kemptide but did not decrease activity. This indicates that stability of the C spine is important for optimization of the catalytic process.

Interconnectivity of the conserved spatial structures sheds light on long-range communication within kinase molecules, which has been observed by numerous authors (13, 14, 22, 24). The unphosphorylated apo-structures are typically the most disorganized: both the R and C spines are broken, and movements of the lobes are not coordinated. R spine assembly induced by phosphorylation or interaction with other activating entities such as cyclin for cyclin-dependent kinases causes significant ordering of the kinase core structure (11, 12). ATP binding completes the C spine formation and makes the molecule even more compact and primed for catalysis. Finally, substrate binding connects all parts of the molecule. The phosphorylation site is in the central cross-point for all conserved structures, and residues positioned close to it can substantially influence the intramolecular connectivity and thus, communication. A well examined case of such influence is $Y^{204}$ from the YLAPEL-motif. Multiple studies demonstrated that mutation of this tyrosine to alanine decreased substrate binding and enzyme activity (22), caused destabilization in distal parts of the C lobe (23), and disrupted general intermolecular communication (24).

## Methods

**Modification of the Computational Method.** LSP alignment is a graph-theory-based method that compares two protein structures and detects similar spatial patterns made by residues in the proteins. The patterns are described by a pair of isomorphic graphs, where vertices correspond to the detected residues with edges connecting pairs of similar residues whose mutual orientation in space is conserved. Our previous work showed that functionally important residues are usually positioned in the middle of the graph with numerous connections to their neighbors. Alternatively, residues that play a supportive role are on a periphery of the graph with a fewer number of connections. The number of connections on the similarity graph was termed IS (see *SI Text* for detailed explanation of the involvement score concept). IS calculation in the current work was made according to the previously published algorithm (8) with a certain modification of similarity specification for residues. Earlier, we considered residues to be similar if their BLOSUM62 coefficient (32) was $\geq 2$. However, this approach sometimes leads to suggestions that are not adequately justified. For example, valine and methionine are considered similar to isoleucine but not to leucine. A decrease of the threshold to 1 resolves this problem but, on the other hand, provides a rather doubtful suggestion that threonine can be substituted by proline, glycine, or aspartic acid. In the current work, we based our substitution matrix on one of the optimized substitution matrices (33) with consideration of the strong cysteine hydrophobicity (34) (see Table S4).

**Protein Structures.** The following structures of protein kinases were used in the analysis: AGC kinases: PKA–PDBID: 2CPK; PKC–PDBID: 2JED; ROCK1–PDBID: 2ESM. CMGC kinases: CDK2–PDBID: 1FIN; SKY1–PDBID:1HOW. Calcium/calmodulin kinases: PHK–PDBID: 2PHK; DAPK–PDBID: 1JKK. Tyrosine kinases: IRK–PDBID: 1IR3; SRC–PDBID 1Y57.

Molecular graphics were prepared by using PyMOL (DeLano Scientific). Molecular surface was rendered with a probe radius of 1.4 Å.

**ACKNOWLEDGMENTS.** This work was supported by National Institute of General Medical Sciences Grant GM70996 (to L.F.T.E.) and National Institutes of Health Grant GM19301 (to S.S.T.).

1. Manning G, Plowman GD, Hunter T, Sudarsanam S (2002) Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci* 27:514–520.
2. Ortutay C, Valiaho J, Stenberg K, Vihinen M (2005) KinMutBase: A registry of disease-causing mutations in protein kinase domains. *Hum Mutat* 25:435–442.
3. Hanks SK, Hunter T (1995) Protein kinases 6. The eukaryotic protein kinase superfamily: Kinase (catalytic) domain structure and classification. *FASEB J* 9:576–596.
4. Huse M, Kuriyan J (2002) The conformational plasticity of protein kinases. *Cell* 109:275–282.
5. Johnson DA, Akamine P, Radzio-Andzelm E, Madhusudan M, Taylor SS (2001) Dynamics of cAMP-dependent protein kinase. *Chem Rev* 101:2243–2270.
6. Taylor SS, Ghosh G (2006) Protein kinases: Catalysis and regulation. *Curr Opin Struct Biol* 16:665–667.
7. Berman HM, *et al.* (2005) The cAMP binding domain: An ancient signaling module. *Proc Natl Acad Sci USA* 102:45–50.
8. Kornev AP, Haste NM, Taylor SS, Ten Eyck LF (2006) Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc Natl Acad Sci USA* 103:17783–17788.
9. Kornev AP, Taylor SS, Ten Eyck LF (2008) A generalized allosteric mechanism for cis-regulated cyclic nucleotide binding domains. *PLoS Comput Biol* 4:e1000056.
10. Iyer GH, Moore MJ, Taylor SS (2005) Consequences of lysine 72 mutation on the phosphorylation and activation state of cAMP-dependent kinase. *J Biol Chem* 280:8800–8807.
11. Lee T, Hoofnagle AN, Resing KA, Ahn NG (2005) Hydrogen exchange solvent protection by an ATP analogue reveals conformational changes in ERK2 upon activation. *J Mol Biol* 353:600–612.
12. Barrett CP, Noble ME (2005) Molecular motions of human cyclin-dependent kinase 2. *J Biol Chem* 280:13993–14005.
13. Emrick MA, *et al.* (2006) The gatekeeper residue controls autoactivation of ERK2 via a pathway of intramolecular connectivity. *Proc Natl Acad Sci USA* 103:18101–18106.
14. Bukhtiyarova M, Karpusas M, Northrop K, Namboodiri HV, Springman EB (2007) Mutagenesis of p38alpha MAP kinase establishes key roles of Phe169 in function and structural dynamics and reveals a novel DFG-OUT state. *Biochemistry* 46:5687–5696.
15. Suel GM, Lockless SW, Wall MA, Ranganathan R (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol* 10:59–69.
16. Swain JF, Gierasch LM (2006) The changing landscape of protein allostery. *Curr Opin Struct Biol* 16:102–108.
17. Kannan N, Neuwald AF (2005) Did protein kinase regulatory mechanisms evolve through elaboration of a simple structural component? *J Mol Biol* 351:956–972.
18. Burk DL, Hon WC, Leung AK, Berghuis AM (2001) Structural analyses of nucleotide binding to an aminoglycoside phosphotransferase. *Biochemistry* 40:8756–8764.
19. Blatch GL, Lassle M (1999) The tetratricopeptide repeat: A structural motif mediating protein–protein interactions. *BioEssays* 21:932–939.
20. Valiev M, Yang J, Adams JA, Taylor SS, Weare JH (2007) Phosphorylation reaction in cAPK protein kinase-free energy quantum mechanical/molecular mechanics simulations. *J Phys Chem B* 111:13455–13464.
21. Scheeff ED, Bourne PE (2005) Structural evolution of the protein kinase-like superfamily. *PLoS Comput Biol* 1:e49.
22. Moore MJ, Adams JA, Taylor SS (2003) Structural basis for peptide binding in protein kinase A. Role of glutamic acid 203 and tyrosine 204 in the peptide-positioning loop. *J Biol Chem* 278:10613–10618.
23. Yang J, *et al.* (2005) Allosteric network of cAMP-dependent protein kinase revealed by mutation of Tyr204 in the P+1 loop. *J Mol Biol* 346:191–201.
24. Masterson LR, Mascioni A, Traaseth NJ, Taylor SS, Veglia G (2008) Allosteric cooperativity in protein kinase A. *Proc Natl Acad Sci USA* 105:506–511.
25. Zhu G, *et al.* (2005) A single pair of acidic residues in the kinase major groove mediates strong substrate preference for P-2 or P-5 arginine in the AGC, CAMK, and STE kinase families. *J Biol Chem* 280:36372–36379.
26. Pike AC, *et al.* (2008) Activation segment dimerization: A mechanism for kinase autophosphorylation of non-consensus sites. *EMBO J* 27:704–714.
27. Kim C, Cheng CY, Saldanha SA, Taylor SS (2007) PKA-I holoenzyme structure reveals a mechanism for cAMP-dependent activation. *Cell* 130:1032–1043.
28. Wu J, Brown SH, von Daake S, Taylor SS (2007) PKA type II-alpha holoenzyme reveals a combinatorial strategy for isoform diversity. *Science* 318:274–279.
29. Song H, *et al.* (2001) Phosphoprotein-protein interactions revealed by the crystal structure of kinase-associated phosphatase in complex with phosphoCDK2. *Mol Cell* 7:615–626.
30. Wu J, *et al.* (2005) Crystal structure of the E230Q mutant of cAMP-dependent protein kinase reveals an unexpected apoenzyme conformation and an extended N-terminal A helix. *Protein Sci* 14:2871–2879.
31. Batkin M, Schvartz I, Shaltiel S (2000) Snapping of the carboxyl terminal tail of the catalytic subunit of PKA onto its core: Characterization of the sites by mutagenesis. *Biochemistry* 39:5366–5373.
32. Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA* 89:10915–10919.
33. Hourai Y, Akutsu T, Akiyama Y (2004) Optimizing substitution matrices by separating score distributions. *Bioinformatics* 20:863–873.
34. Nagano N, Ota M, Nishikawa K (1999) Strong hydrophobic nature of cysteine residues in proteins. *FEBS Lett* 458:69–71.